

SIMILARITY SEARCHING OF MOLECULES BASED UPON STATISTICAL ANALYSIS OF DESCRIPTOR VECTORS CHARACTERIZING MOLECULAR REGIONS

Abstract of the Disclosure

5 The method of the present invention provides for similarity searching of molecules based upon statistical analysis of descriptor vectors characterizing molecular regions. In a training phase, an association criterion is generated by which query regions of a query molecule are associated with regions of molecules stored in a database. Preferably, the association criterion is based upon statistical analysis of groups of descriptor vectors that characterize properties of the regions of the molecules stores in the database. In an acquisition phase, for each molecule in a series of molecules, the following steps are performed for a given molecule. Data that represents the structure of the given molecule is read from persistent memory and used to define a set of three-dimensional regions of space in the given molecule. For each region, one or properties of the given molecule are mapped to property values for grid points of the region. A multi-map entry is generated that identifies the region, and position and orientation of set of axes derived from the property values of the grid points of the region. The association criterion generated in the training phase is used generate a key for the region, and the entry is stored in the multi-map at a location associated with the key. In the recognition phase, data that represents the structure of a query molecule is used to define a set of regions in the query molecule. For each region, one or properties of the query molecule are mapped to property values for grid points of the query region. The association criterion generated in the training phase is used generate a key for the query region. The multi-map entry identified by the key is retrieved and the data stored therein are read from the table. For each stored region identified by the retrieved table entry, an hypothesized match is constructed and added to a vote table. After processing all of the stored regions identified by the retrieved multi-map entry for the set of query regions in the query molecule, one or more entries of the vote table is selected, the alignment transformations stored in the selected entries are applied to corresponding molecules stored in the database, and the

